Transmission

12

Learning Objectives

After completing this reading, you should be able to

- Describe the role of a transmission system and explain the challenges related to electricity transmission
- Compare and explain how different factors can determine the capacity of a transmission line, including thermal limits, voltage limits, and angle stability limits
- Describe active and reactive power and understand their role in voltage control
- Apply binding constraints to determine the capacity of a transmission line and describe considerations in determining a safety margin for each type of limit
- Describe Kirchhoff's Laws and use these laws to calculate the feasibility of dispatching power on each line in a hypothetical system with two generators and two transmission lines
- Describe the role of ancillary services and compare characteristics of different types of ancillary services
- Define losses in an electrical system and explain different ways in which losses can arise

Excerpt is Chapter 4 of Introduction to Electricity Markets, *by John Parsons.*

Chapter 4

Transmission

4.1 Introduction

This chapter presents a few key fundamentals about the transmission system that are especially important to electricity markets. The transmission system carries electric power over long distances, connecting many generating units to many distribution networks with load. It consists of (i) high voltage power lines, usually strung overhead on high towers, but sometimes buried underground, (ii) transformers, used to change the level of voltage, which allows different parts of the system to connect with one another at the right voltages, (iii) switchgear, such as circuit breakers, that can disconnect parts of the network to enable maintenance or to protect the system, and (iv) measurement instrumentation that collects voltage, current and power data for monitoring, control and metering purposes, and communications equipment that reports these data to control centers and allows switchgear to be controlled remotely. The transformers, switchgear, measurement and communications equipment are usually housed in stations and substations located throughout the network.

Our main concern is not with the operation of the transmission equipment itself, but with how the transmission network connects generators and load into one large electricity system that must be managed as one unit. Reliability requires that all parts of the system–from the generators at one end, to the load at the other, and all of the equipment in between–be operated in tight coordination as a single machine. Total generation on the network must be continuously and precisely matched with total load. Voltages throughout the network must be kept within tight bounds. Power flows must be planned to assure stability. A failure on these dimensions can lead to costly outages, which, in extreme cases, can cascade and lead to a large-scale blackout.

Assuring the security and stability of the system requires planning for the inevitable disturbances and contingencies, small and large, brief or lasting. A system can include thousands of generators located across a large geographic territory, millions of consumers accessing the system to power an enormous array of equipment, connected by a meshed network of transmission lines and associated transmission equipment. Generators sometimes suddenly go offline. Transmission lines are occasionally tripped and go briefly out-of-service. When that happens, power flows move instantaneously onto other lines. Transformers and other equipment may fail, or switches and relays may malfunction. The system must quickly adjust to these and other contingencies, either automatically or at the direction of the system operator. Without proper controls and adjustments, system frequencies and voltages at certain locations may move outside of safe bounds. As a protective measure, many parts of the system are designed to automatically disconnect themselves under certain circumstances. The system operator can also purposefully disconnect or de-energize portions of the network, using sensing equipment, circuit breakers and switches as needed. While this protects the equipment and people, it can also exacerbate the instability: an **outage** at one generator can create circumstances that trigger another generator to trip and be disconnected, which may only exaggerate the problem, potentially leading other generators to trip and disconnect, and so on. An overload of power on a line that faults and disconnects, sends the excess power flowing onto other lines which themselves may not be able to handle the

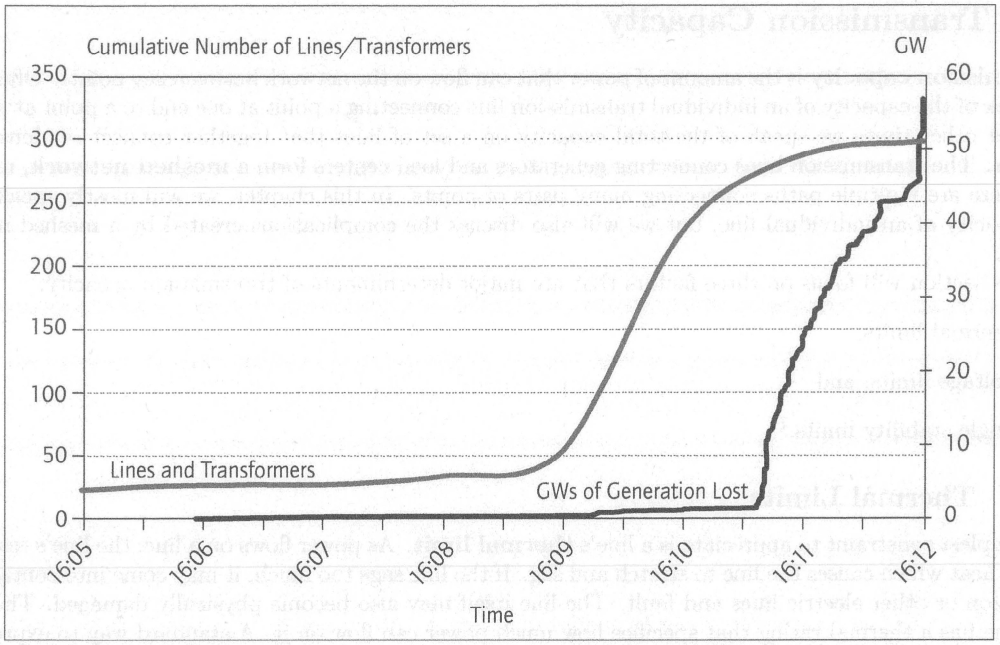


Figure 4.1: Line Trips and Lost Load During the Cascade Phase of the 2003 Blackout in North America.

Image credit: US-Canada Power System Outage Task Force, Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations, 2004.

surge, and so on. Events like these can lead to a **cascade** of outages or a **blackout**.

Figure 4.1 shows the cascade that struck parts of North America in August, 2003 following the tripping of some transmission lines in the state of Ohio. Failure of an additional important line occurred at 16:06 (4:06 pm). As the graph shows, by 16:09 (4:09 pm) or less than three minutes later, a cascade of outages of hundreds of transmission lines and transformers occurred. This was quickly followed by a cascade of generators tripping off-line, amounting to more than 50 GW of generation lost. In the end, more than 50 million customers lost power. Major blackouts are rare but costly events. In addition to the North American blackout of 2003, some famously instructive ones are Tokyo 1987, Brazil 1999, Switzerland/Italy 2003, and Australia 2004. System operators use these cases to identify better ways to anticipate problems and better ways to design systems to avoid damage.

Operation of the system is extremely complex and requires extensive knowledge of the physical principles underlying electricity as well as the engineering properties of a wide assortment of equipment. In this chapter we introduce just a few of these principles and properties of the transmission network that are most important to how generation is managed and how electricity is traded and priced. We organize the rest of this chapter around three topics:

- transmission capacity,
- ancillary services, and
- losses.

4.2 Transmission Capacity

Transmission capacity is the amount of power that can flow on the network between key points. Oftentimes we speak of the capacity of an individual transmission line connecting a point at one end to a point at another end, but other times we speak of the total capacity on a set of lines that together connect sections of the network. The transmission lines connecting generators and load centers form a **meshed network**, meaning that there are multiple paths connecting many pairs of points. In this chapter, we will mostly speak about the capacity of an individual line, but we will also discuss the complications created by a meshed network of lines.

This section will focus on three factors that are major determinants of transmission capacity:

- thermal limits,
- voltage limits, and
- angle stability limits.[1]

4.2.1 Thermal Limits

The simplest constraint to appreciate is a line's **thermal limit**. As power flows on a line, the line's resistance creates heat which causes the line to stretch and sag. If the line sags too much, it may come into contact with vegetation or other electric lines and fault. The line itself may also become physically damaged. Therefore, each line has a thermal rating that specifies how much power can flow on it. A standard way to express this rating is in terms of the continuous flow that is allowed. The rating is established based on assumptions about the conditions in which the line is operating. A transmission line will stretch and sag more on a hot summer day with little wind than on a cold winter day with significant wind. Therefore, the thermal limit may be established on the basis of a worst case assumption about external conditions.

There may be multiple ratings for a line. For example, there may be an emergency rating that allows a greater power flow in exceptional conditions, but for a limited duration. A line can also have a dynamic rating, taking into account external conditions like the season or ambient temperature and other factors can affect the line's ability to handle the production of heat from the power transfer. A dynamic rating increases the line's capacity in most circumstances.

4.2.2 Voltage Limits

Another constraint is the **voltage limit**. Voltages in a system must be kept within strict tolerances. Voltages that are too high or too low can damage equipment. Extremely high voltages can cause dangerous arc flashovers. Low voltages can cascade into a voltage collapse. A typical tolerance is ±5% of the system's **nominal voltage**, which is the level specified for normal operation. For example, if the system is designed with a nominal voltage of 400 kV, then 5% is 20 kV, so that the upper limit of 105% is 420 kV and the lower limit of 95% is 380 kV.

A major cause of low voltage is a lack of **reactive power**. Appreciating the nature of this problem requires understanding an important complexity of alternating current (AC) power flow. Although we often speak as if there is only one type of power, in fact, an AC system has two components of power flow.

- Active power – First, there is the component known as **active power**, which is the power that performs useful work in electrical devices, generating heat in a toaster, lighting a light, activating the motor in a refrigerator or running industrial equipment. Active power is measured in watts (W), kilowatts (kW) or megawatts (MW).[2]

---

[1] Angle stability limits are sometimes called power stability limits. They are also sometimes called transient stability limits, which is confusing since voltage limits may also incorporate transient stability considerations.

[2] Active power is also referred to by the term real power, although we will not use this term.

- Reactive power – Second, there is the component known as reactive power, which is the power used to produce the electric and magnetic fields in electric motors, on transmission lines and other equipment. Reactive power is measured in volt-amperes reactive (var), kilovolt-amperes reactive (kvar) or megavolt-amperes reactive (Mvar).

Taken together, as a pair, the combination of active and reactive power are known as complex power. It is measured in volt-amperes (VA). An index of the combination of active and reactive power is known as apparent power, which is also the magnitude of complex power. The relationships between active power, reactive power and apparent power are critical to operating an electric system. However, in this text we cannot dive fully into these. Instead, we focus on a few important rules of thumb about active and reactive power that help describe key economic features of managing an electric system.

Reactive power is central to AC systems because the systems are built using electric and magnetic fields, and reactive power is the building block for those fields. Both active power and reactive power are produced by generators. Both flow on transmission lines. Both are used by load. In this sense, we can think of the electricity system as producing and distributing two products simultaneously. However, the ways in which the two types of power are produced and flow are very, very different. Active power is what performs the useful work for consumers, such as turning the electrical motor in a refrigerator. Accordingly, active power is what we meter and bill consumers for. In contrast, reactive power is not consumed, but oscillates. It flows into equipment and then flows back out. When an electric motor starts up, it needs a large injection of reactive power to create its magnetic field.[3] When the motor stops and the magnetic field is closed, reactive power flows back from the motor into the system. Transmission lines, too, sometimes absorb reactive power, and other times they return that reactive power to the system, depending upon their operating state. On average, the flow of reactive power is zero, but in any window of time the flow may be either positive or negative.

When power is transmitted, some of both types of power are dissipated in the transmission system as losses, which causes the system voltage to decline. In most high voltage, long-distance transmission systems, the losses in reactive power are much more significant. Active power losses on a line are determined by the line's **resistance**, which is measured in ohms (Ω, the Greek letter omega). Resistance is a measure of the opposition to the flow of current in a circuit. Current flows more easily through some materials than others. Copper is a popular material for electrical wiring because current flows very easily through it, since copper has low resistance. Reactive power losses on a line are determined by the line's **reactance**, which is also measured in ohms (Ω). Reactance is a measure of the opposition to the change in current or voltage in a circuit. A high voltage transmission line's reactance is typically much greater than its resistance–a typical 100-mile long 345 kV line may have a resistance of 6 Ω and reactance of 60 Ω.[4] Consequently, the reactive power losses are much greater. In fact, these reactive losses are so great that it is not possible to deliver significant quantities of reactive power over long distances. In order to maintain the system voltage along the entire transmission system, it is necessary to inject reactive power locally.

The voltage drop due to losses from reactive power is often represented in a P-V curve. A sample P-V curve is shown in Figure 4.2. The horizontal axis is the rate of power transfer over a transmission line. The vertical axis is the voltage at the receiving substation. The system's nominal voltage is 350kV, which should be the voltage at the sending end of the line. For small amounts of power transfer, the voltage drop from 350kV is relatively small. However, as the power transfer increases, the rate of voltage drop begins to increase. Too heavy a power transfer results in high reactive power losses and a dangerous drop in voltage. At a certain point, the drop is dramatic and the system becomes unstable. In this diagram that point of instability–known as either the 'knee' or the 'nose' of the P-V curve–occurs as the power transfer approaches

---

[3] A familiar example of voltage drop occurs in many households when a refrigerator motor kicks on and lights briefly dim. Operation of the refrigerator motor requires a sudden injection of reactive power to create a magnetic field, and this causes a voltage drop in the house, which is experienced as a decline in the power flowing through the lights. In an instant, once the magnetic field has been created, the need for reactive power drops and the system quickly restabilizes back to its normal voltage level and the power flowing through the lights returns to its normal level. While this example may be familiar, it is a minor contributor to reactive power needs on a transmission system.

[4] EPRI Power Systems Dynamics Tutorial, July 2009, Project Manager Guorui Zhang, p. 5-7. The exposition above elides the distinction between inductive reactance and capacitative reactance.

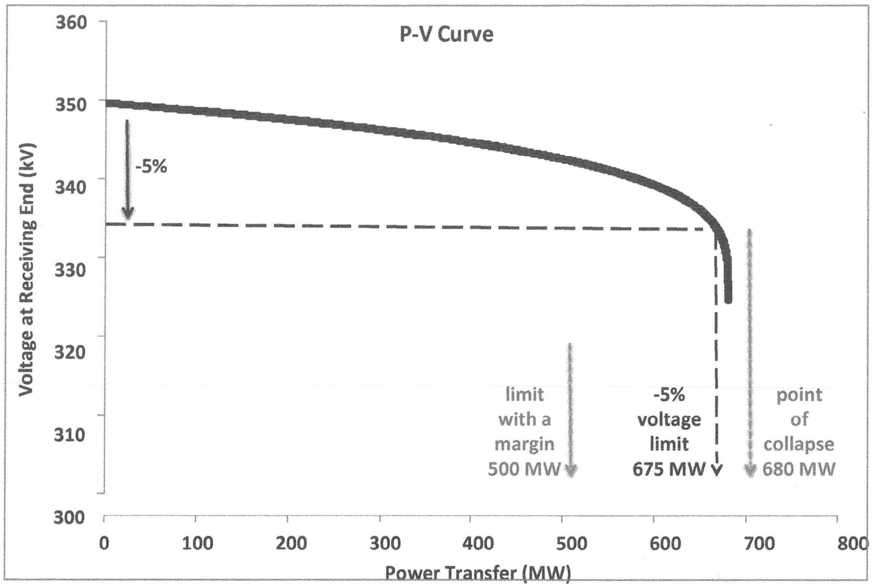


Figure 4.2: A sample P-V curve. As Power Transfer Across the Line Approaches 700 MW, a Voltage Collapse Occurs.

680 MW. At this point a voltage collapse may begin.

The transmission system operator can protect against the voltage drop by injecting reactive power at the receiving end of the line. This can be done from generators or other equipment such as capacitors, reactors and tap changing transformers. However, at any point in time, there is a limit to the available reactive power. Once this limit is reached, it is important to avoid a damaging loss of voltage by limiting the amount of power flowing on the line. Voltage limits cap the amount of power transfer permitted across the line.

4.2.3 Angle Stability Limits

The third important constraint is the **angle stability limit**. These are also known as power flow stability limits and sometimes as transient stability limits. Angle stability limits are necessary to assure that generation and loads remain synchronized. A loss of synchronization can damage equipment. A system in which various groups of generation and load are losing synchronization can eventually break into separate electrical islands.

The importance of angle stability limits stems from another feature of AC power flow. Voltage at the load end of the line is out of phase with the voltage at the generator end. An increase in load and power flow down a line increases the phase difference between the voltage at the generator end and the voltage at the load end. There is a limit to this difference. A phase difference of 90 degrees corresponds to the theoretical maximum power flow. Transfers beyond this limit can produce a voltage collapse, albeit for a different reason than the lack of reactive power discussed above. In order to maintain transient stability–stability in the face of a contingency–the power angle limit is set below 90 degrees. Limits in the neighborhood of 30 and 45 degrees are commonly applied.[5] This angle stability limit can then be translated into a limit on the power flow on the line.

---

[5] IEA, 2005, Learning from the Blackouts, p. 30.

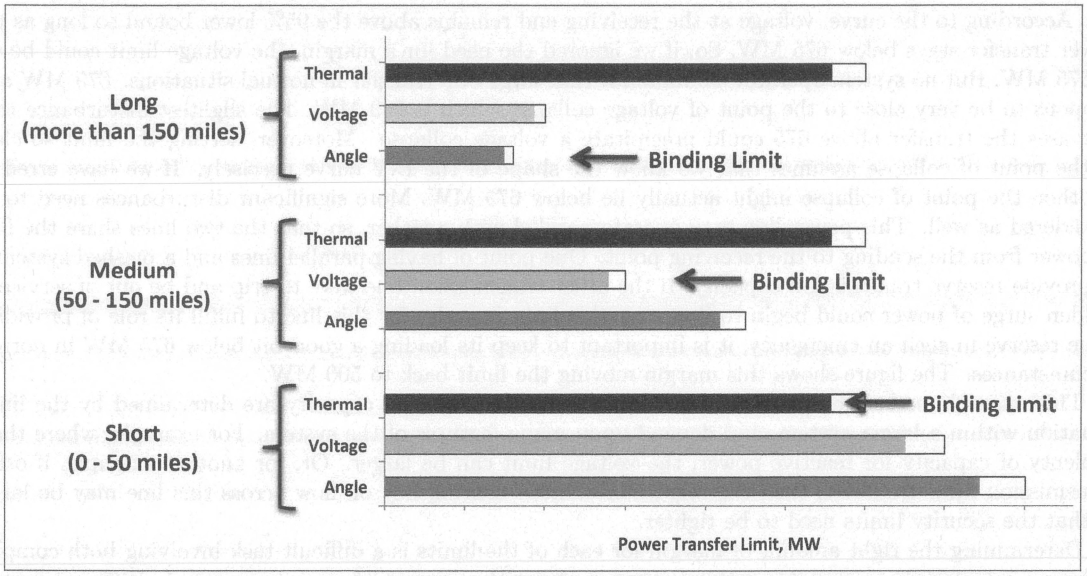


Figure 4.3: Determining a Binding Limit on Power Flows Across Lines of Different Lengths

4.2.4 Determining Capacity

The capacity of a transmission line will be determined by whichever of the three constraints is binding. Figure 4.3 illustrates this for three different lengths of a line: short, medium, and long.[6] For each length, it shows the three constraints, thermal, voltage and angle stability. For each length, one of the three is the binding constraint: whichever provides the lowest cap on power transfer. The thermal limit of a line is independent of the length of the line. That is why Figure 4.3 shows the same thermal limit for all three line lengths. Voltage and angle stability limits generally become tighter with the length of the line, just as shown in Figure 4.3. That is why voltage and angle stability limits are more likely to be the binding constraint on longer lines while thermal limits are usually the binding constraint on short lines, as shown in Figure 4.3. But this is a generalization and the actual ranking of the constraints depends on many factors besides the length of the line.

In order to ensure the stability and security of the system, these limits often need to incorporate a **safety margin**. They are set not only so that the system can operate as planned, but also so that the system remains stable and secure in the face of disturbances. These include the normal fluctuations in load or generators, as well as major **contingencies** such as a transmission line outage, a major generator suddenly going off-line. Operators often distinguish these according to the time horizon of the disturbance and response. Some disturbances must be resolved in milliseconds using automatic devices and processes that quickly restore the system to its normal operating condition. In this chapter, we use the term **stability** for this property–a quick and automatic return to normal operation. Other disturbances must be resolved over a longer window of time, and can involve operator decisions about how to respond, including which generators to re-dispatch or how to alter the transmission system in one way or another to restore the system to its normal operating condition. In the meantime, the system operating parameters must stay within acceptable bounds. We use the term **security** for this property–the ability to deliberately return the system to normal operation within a slightly longer, defined window of time, while staying within defined bounds in the meantime. Including safety margins helps to assure both stability and security. Each type of limit–thermal, voltage and angle stability–will need some margin.

We can illustrate the application of a margin using voltage limits and the P-V curve shown in Figure

---

[6] This is a reproduction of Figure 2.4 from the MIT Future of the Electric Grid study, 2003, with certain relabeling.

4.2. According to the curve, voltage at the receiving end remains above the 95% lower bound so long as the power transfer stays below 675 MW. So, if we ignored the need for a margin, the voltage limit could be set at 675 MW. But no system operator would allow that large of a transfer in normal situations. 675 MW also happens to be very close to the point of voltage collapse which is 680 MW. The slightest disturbance that increases the transfer above 675 could precipitate a voltage collapse. Moreover, setting the limit so close to the point of collapse assumes that we know the shape of the P-V curve precisely. If we have erred at all, then the point of collapse might actually lie below 675 MW. More significant disturbances need to be considered as well. This power line may operate parallel with another, so that the two lines share the flow of power from the sending to the receiving point. One point of having parallel lines and a meshed system is to provide reserve transmission capacity. If the other transmission line were to trip and be out of service, a sudden surge of power could begin to flow over this line. In order for this line to fulfill its role of providing some reserve in such an emergency, it is important to keep its loading a good bit below 675 MW in normal circumstances. The figure shows this margin moving the limit back to 500 MW.

This example makes the point that the limits on a line's normal capacity are determined by the line's situation within a larger system, and depend upon many features of the system. For example, where there is plenty of capacity for reactive power, the voltage limit can be larger. Or, for another example, if other transmission lines are down, then the impact of another contingency on flow across this line may be large, so that the security limits need to be tighter.

Determining the right amount of margin for each of the limits is a difficult task involving both complex technical calculations as well as a value judgment about the amount of security required. Different system operators will formalize the determination of security limits in different ways. One example of this formalization is known as the **"N-1 criterion"** which says that the system must remain secure–i.e., capable of maintaining frequency, current and voltages within emergency ratings– following the unexpected loss of the most important generator or transmission facility. In addition, the system must be restored to normal limits within some specified period of time, to a point where it can handle the loss of the next most important generator or transmission facility. The size of the contingencies on a system will determine the size of the margins needed and therefore how tight are the limits. A small system with a couple of very, very large thermal generators needs larger margins–tighter limits–because the outage of a single generator will have a large impact. In practice, security margins are often more finely tuned than this simple criterion suggests. In order to provide an extra margin of safety for densely populated cities, for example, the grid may be operated to withstand the concurrent loss of two or more facilities instead of just one. Or, where system operators assess that an event damaging one structure is highly likely to also damage another, then the system will need to be resilient to the simultaneous loss of both.

The operator also needs to account for the possibility that power flows differently than the operator's model predicts. It is not possible to have highly precise real-time knowledge of network voltages and impedances throughout the system, and so it is not possible to predict power flows on a given transmission line with precision. Moreover, most systems are really a part of a larger network, and the flows from that larger network may change the flows on the system being analyzed. These unplanned flows are known as **loop flows**. Some allowance must be made for this uncertainty, too. Better monitoring and modeling may be able to reduce this uncertainty. This can reduce the margins required, relax the limits, and effectively expand transmission capacity.

Finally, the size of the margins required will also depend on the ability of the system operator to respond to a contingency and quickly restore the system to normal operation. A key ingredient is preparation for contingencies. A well-managed operator will be regularly running computer programs to project system power flows and voltages under different contingencies. This will enable it to quickly determine the best course of action when an actual problem arises. The ability to respond will also depend on having in place systems for swift communication and intervention at operationally or strategically important locations in the system. The more swift and assured is the response, the less restrictive can be the limits imposed beforehand.

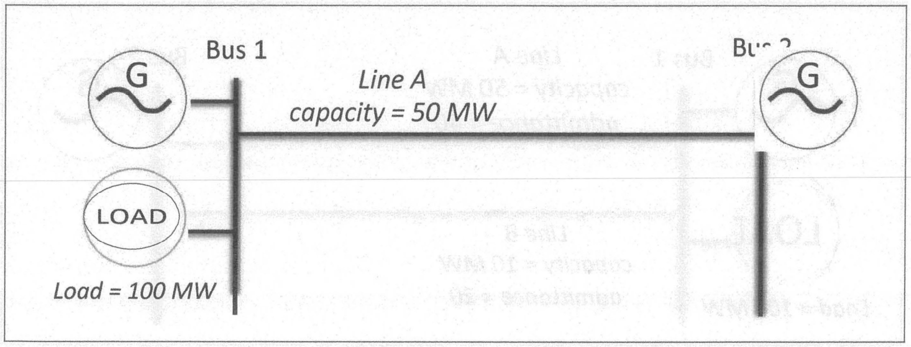


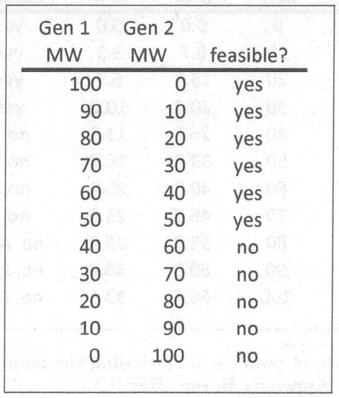
Figure 4.4: Load and Generation at Two Buses Connected by a Single Transmission Line.

4.2.5 Transmission Capacity and Generation Dispatch

In a simple radial system, the impact of a transmission constraint is straightforward. Figure 4.4 shows such a system. There are two buses, one at each end, connected by a single transmission line, Line A. At Bus 1, there is a generator and also a load. At Bus 2, there is another generator. In a diagram like this, we assume that any generation and load connected at the same bus, as at Bus 1, can be matched without any losses or other concerns about the power flow. For simplicity, we have ignored the losses in this example, as we will in a number of examples. We will bring them back into the dispatch decision at a later time. Our only concern here is the amount of power that can flow between the two buses across Line A. The figure shows Line A's maximum capacity. Table 4.1 shows the different possible combinations of generation from the two units. The last column reports whether the combination is feasible with respect to the transmission line's capacity. Any combination in which Generator 2 is scheduled to produce more than 50 MW is infeasible because no more than 50 MW can be transmitted across Line A.

Large scale transmission systems are meshed, and power flows by multiple paths. This complicates the problem of determining the feasible profiles of generation because of the laws of physics that govern the flows of power across the different paths. According to Kirchhoff's Laws, the flow of power is distributed in proportion to each path's **admittance**, which is a physical property of the lines and other equipment creating the path. For example, if two lines are otherwise similarly constructed, the line with the greater cross section will have a greater admittance. A comparable line that travels a shorter distance to the same

Table 4.1: Feasible and Infeasible Generator Dispatch for the Two Buses Connected by a Single Transmission Line.



Gen 1 MW	Gen 2 MW	feasible?
100	0	yes
90	10	yes
80	20	yes
70	30	yes
60	40	yes
50	50	yes
40	60	no
30	70	no
20	80	no
10	90	no
0	100	no

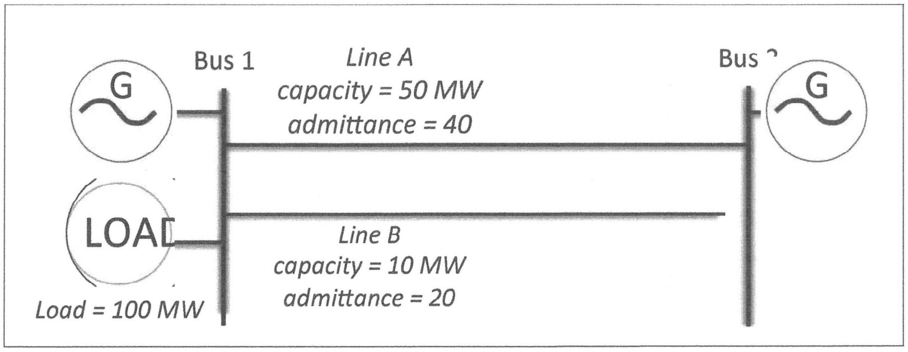


Figure 4.5: Load and Generation at Two Buses Connected by Parallel Transmission Lines.

destination will have a greater admittance.[7]

Figure 4.5 shows a variation on Figure 4.4 in which a second, parallel transmission line has been added connecting the two buses. The figure shows the different admittances of the two lines. Power from Generator 2 will flow to the load in proportion to the two admittances. Writing P_A for the power flow on line A, P_B for the flow on line B, and P_T for the total flow, we have:

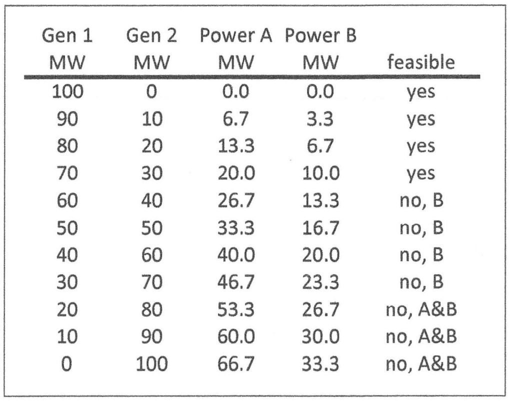
$$P_A = \frac{\gamma_A}{\gamma_A + \gamma_B} P_T = \frac{40}{40 + 20} P_T = \frac{2}{3} P_T \tag{4.1}$$

$$P_B = \frac{\gamma_B}{\gamma_A + \gamma_B} P_T = \frac{20}{40 + 20} P_T = \frac{1}{3} P_T. \tag{4.2}$$

Since the admittance of Line A is twice the admittance of Line B, 2/3 of the total power will flow along line A at the top, and 1/3 will flow along line B at the bottom.

Table 4.2 shows the different possible combinations of generation from the two units, together with the associated flow of power across the two different lines. The last column of the table reports whether the dispatch is feasible. Feasibility requires that the flow on Line A be less than Line A's capacity, and also that the flow on Line B be less than Line B's capacity. The column also reports which line is the reason the

Table 4.2: Feasible and Infeasible Generator Dispatch for the Two Buses Connected by Parallel Transmission Lines.



Gen 1 MW	Gen 2 MW	Power A MW	Power B MW	feasible
100	0	0.0	0.0	yes
90	10	6.7	3.3	yes
80	20	13.3	6.7	yes
70	30	20.0	10.0	yes
60	40	26.7	13.3	no, B
50	50	33.3	16.7	no, B
40	60	40.0	20.0	no, B
30	70	46.7	23.3	no, B
20	80	53.3	26.7	no, A&B
10	90	60.0	30.0	no, A&B
0	100	66.7	33.3	no, A&B

---

[7] Other factors enter into the determination of power flow, including the amplitude and phase of the voltages at each end. See the MIT Future of the Grid study, 2011, Appendix B, esp. Box B.1.

Table 4.3: Revised Feasible and Infeasible Generator Dispatch for the Two Buses Connected by Parallel Transmission Lines, Given Lower Admittance on Line B.

Gen 1 MW	Gen 2 MW	Power A MW	Power B MW	feasible
100	0	0.0	0.0	yes
90	10	8.3	1.7	yes
80	20	16.7	3.3	yes
70	30	25.0	5.0	yes
60	40	33.3	6.7	yes
50	50	41.7	8.3	yes
40	60	50.0	10.0	yes
30	70	58.3	11.7	no, A&B
20	80	66.7	13.3	no, A&B
10	90	75.0	15.0	no, A&B
0	100	83.3	16.7	no, A&B

dispatch is infeasible. Surprisingly, although we have added new transmission capacity, the maximum we can dispatch from Generator 2 is only 30 MW, which is less than the maximum before Line B was added! This is because the relative admittances of the two lines demands that 1/3 of the power flow on Line B, and Line B's capacity is less than 1/3 of Line A's capacity. It is not physically possible to apportion the flow differently given the system as designed here.[8] Therefore, Line B's limitation becomes a limitation for the system as a whole.

This example illustrates the importance of Kirchhoff's Laws for our understanding of systemwide capacity. It also illustrates how interdependent all parts of the system are upon all other parts of the system. This turns out to be a common feature of electricity transmission systems, for this as well as some other reasons.

Of course, if we had intended the added transmission line to increase the total transfer capacity from Bus 2 to Bus 1, then we would have designed it with these characteristics in mind. We would have either built a line with greater capacity, or we would have built a line with a lower admittance so that a smaller fraction of power flowed over Line B, a fraction more consonant with the line's capacity. For example, suppose that Line B had an admittance of 8 instead of 10. Table 4.3 shows how this changes the associated flow of power across the two different lines for each combination of generation. In this case, adding Line B has expanded the feasible dispatch from Generation 2: now it is feasible to use Generation 2 for up to 60 MW of power.

4.3 Ancillary Services

The term **ancillary services** is a catch-all for the many ways in which generators, along with other equipment, can contribute to the stability and security of the system. Different regions and countries define different categories of ancillary services and use different labels. In this text, we focus on three categories:

- frequency control, including regulation and operating reserves,
- voltage support, and
- black start or system restart.

The provision of each service requires investments in equipment and careful planning for how the equipment is operated.

---

[8] Of course, we can disconnect Line B. There are also some other tools that can be used. Together, these tools are known as *flexible AC transmission systems*, or FACTS. These are not often employed in most networks. They are relatively new and expensive, although it is likely that their development and use will advance quickly. In any case, we don't consider them here.

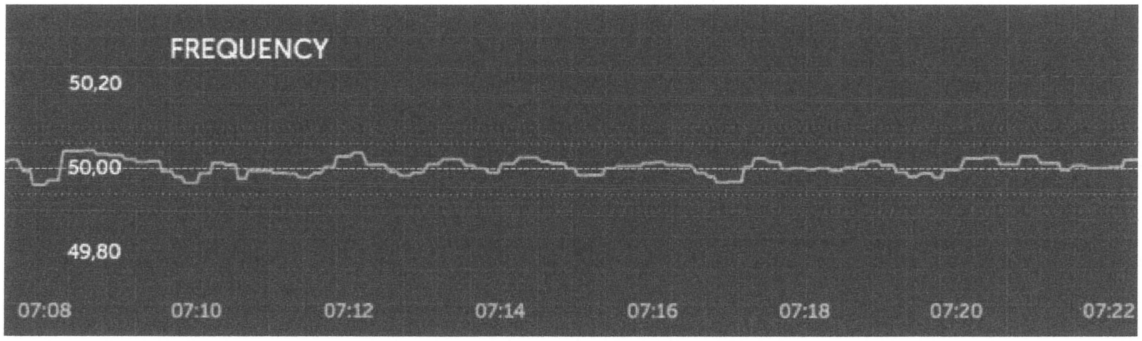


Figure 4.6: Frequency Reported by the Norwegian Grid Operator, Statnett, During a Few Minutes on March 4, 2017.

Image credit: Statnett website. Reprinted with permission.

4.3.1 Frequency Control

A constant **balance** between generation and load is key to operating the system within bounds and avoiding faults. All systems have a certain predictability in the load pattern through the day, and the system operator will produce a plan to dispatch generators in order to serve the load. However, for any number of reasons, actual load will vary from the forecasted load, and actual generation will vary from the dispatch plan. Some of these variances will be very small perturbations that occur in very short time scales, and some may be surprisingly large discrepancies lasting for hours. The system operator must have means at its disposal to quickly adjust generation accordingly and to maintain a constant balance between generation and load in the face of every variance, whether small or large, brief or long lasting.

A key metric for measuring the balance between generation and load is the system's frequency. When generation exactly matches load, the system will be at its design frequency. If load increases without a matched increase in generation, then frequency falls. The system responds by dispatching additional generation, as we shall see, and thereby returns the frequency to its designed level. Figure 4.6 shows the frequency on the Norwegian grid during a few minutes on March 4, 2017 as reported by the grid operator Statnett. The design frequency is 50 Hz. The figure shows the system's frequency varying slightly around this number, but generally staying within the normal operating range of 49.90-50.10 Hz.

Because frequency is the diagnostic, maintaining an instantaneous balance between generation and load is called **frequency control**. Tools for frequency control operate in different ways and with different speeds. Some tools operate instantly and automatically, while others require decisions by the system operator and take longer to implement. Figure 4.7 illustrates the progress of events when there is a grid event causing a drop in frequency. Initially, the system frequency drops dramatically. The rate of drop is immediately dampened by the system's inertial response, and automatic controls begin to increase generation or decrease load to arrest the drop. Further automated actions quickly cause system frequency to rebound. This all happens within seconds. As we move through time, the figure's time dimension switches to minutes. The system operator begins to take charge of the process, increasing the dispatch of specific generators to restore the systemwide balance with load. Gradually the system frequency is restored to its proper level and regular operation can resume.

Here is some additional detail on the different types of frequency control employed:

Inertial Response

Each of the rotating masses of generators and motors connected to the grid possesses inertia. Should there be a sudden change in the generation or load of the system, this inertia automatically helps to stabilize the

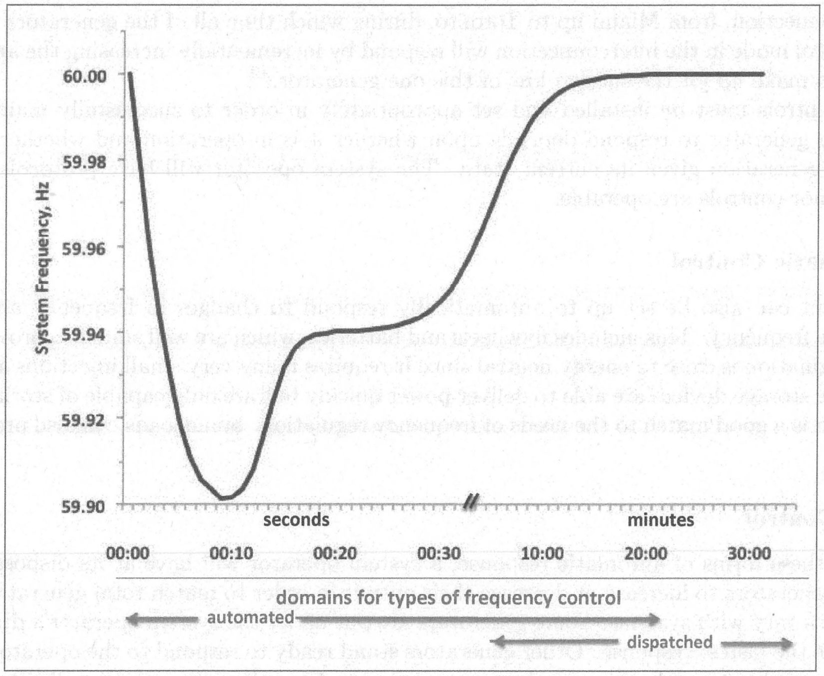


Figure 4.7: Timing of Responses to a Grid Event Causing a Frequency Drop.

system. If a generator on the system unexpectedly reduces the power it is supplying, the frequency of the system begins to fall. The **inertial response** of other generators to this frequency drop is a release of some of their stored kinetic energy, which modulates the initial rate of change of frequency.

The system will have more or less inertial response depending upon the quantity and type of equipment up and running and connected to the system. System inertia is proportional to the sum of stored energy in the rotating masses of generators and motors directly connected to the grid. The amount of inertia each generator contributes to the system depends on the size and type of the generator when it is running. It does not depend directly on how much power it is producing. Even if a generator is part-loaded it contributes to the total system inertia. The system operator can change the dispatch of generators to change the total amount of inertial response on the system. High system inertia makes for a strong system that can handle transient changes in system frequency. Low system inertia increases the likelihood that a succession of small disturbances could lead to severe faults or loss of generation or demand.

Inertial response occurs in milliseconds. It temporarily modulates the system's response to the imbalance between generation and load, which gives the system operator time to remove the imbalance through other means.

Governor Control

A governor can be used to modulate the operation of certain types of generators in response to changes in frequency. The governor controls the amount of steam, water or gas flowing to the turbine. When changes in the frequency are sensed, the governor responds by altering the flow to the turbine. If the frequency has decreased, reflecting a deficiency of generation to load, the governor will increase the flow, thus increasing the generator's output and bringing generation and load back into balance. The governor is set to maintain the local frequency close to the established standard.

Governor control is a form of automatic response that takes place in fractions of a second. In North America, for example,"the sudden loss of a large generator in Kentucky can cause a frequency dip in the entire

Eastern Interconnection, from Miami up to Toronto, during which time all of the generators with automatic generation control mode in the interconnection will respond by incrementally increasing the amount of energy they produce to make up for the sudden loss of this one generator."[9]

Governor controls must be installed and set appropriately in order to successfully maintain frequency. The ability of a generator to respond depends upon whether it is in operation and whether it can increase or decrease its generation given its current state. The system operator will have protocols to ensure that sufficient governor controls are operable.

Other Automatic Control

Other equipment can also be set up to automatically respond to changes in frequency and thereby help quickly stabilize frequency. This includes flywheels and batteries, which are well suited to providing frequency regulation. Regulation is close to energy neutral since it requires many very small injections and withdrawals of power. These storage devices are able to deliver power quickly but are only capable of storing a low volume of energy, which is a good match to the needs of frequency regulation. Some loads can also provide automatic control.

Dispatched Control

In addition to these forms of automatic response, a system operator will have at its disposal the ability to direct certain generators to increase or decrease their output in order to match total generation to load. The protocols for this vary with systems. Some generators are put under the system operator's **direct control** in order to provide the fastest response. Other generators stand ready to respond to the operator's instructions. These units are used to provide the more lasting response and to relieve those under direct control.

In different countries and regions around the world, there are different sets of terms used to distinguish tools that operate at different speeds or with greater or lesser degree of automation. For example, Australia defines a regulation service and three classes of reserves, which it calls contingency services. These classes are distinguished by the response time:

- 6 second – arrest a rapid change in system frequency within the first six seconds of a frequency disturbance, and then provide an orderly transition to the 60 second service.
- 60 second – stabilize the system frequency within the first sixty seconds of a frequency disturbance, and then provide an orderly transition to the 5 minute service.
- 5 minute – restore system frequency to its nominal 50Hz within the first five minutes of a frequency disturbance, and to sustain response until notified by central dispatch.

In the Nordic market, the terminology is:

- Frequency controlled operating reserve (100% activated between 49.9-50.1 Hz)
- Frequency controlled disturbance reserve (50% activated at 5 seconds and 100% at 30 seconds)
- Fast active reserve (15 minutes)
- Slow active reserve (4-8 hours)
- Reactive reserve

The UK distinguishes between frequency response and reserves. There are several types of frequency response:

- Primary
- Secondary
- High
- Low frequency relays for loads

There are four types of reserves:

- Regulating
- Standing
- Fast
- Warming & Hot standby

---

[9] FERC Reliability Primer, 2009, p. 24.

As you can see, each set of terms is organized with respect to automation and speed.[10] Historically, since reserves were generally provided by turbine generators, and those that were up and running already–i.e., spinning–could respond the fastest. Therefore, fast acting reserves were often called spinning reserves, and slower acting reserves were often called non-spinning reserves. However, these days equipment such as batteries, which don't spin, may be included in the first class of reserves, and certain loads can also provide demand response at this speed. So, the terminology has been adapting and making a more generic reference to automation and speed.

In this volume we use the term **regulation** to cover tools acting at the shortest time scales–usually seconds–that generally require automated operation. We use the term **operating reserves** to cover tools acting at slightly longer time scales–usually minutes–requiring human action. We will not make any sharper distinctions.

The amount of regulation and operating reserves required on a system depends on its reliability criteria and on the size of the system. Larger systems generally need a smaller amount of reserves as a fraction of load, than do smaller systems. This is an important motivation for cooperation among neighboring systems. Reserves are often called on because other generators fail, and therefore the amount or type of reserves required depends on the types of generation on a system, how they are maintained and how they are distributed across the transmission system. Systems with a significant amount of intermittent generation, like wind and solar, may require additional regulation. Systems with a few large nuclear stations may require additional operating reserves. The correct amount will depend upon the particular features of the local system.

Nevertheless, for major electricity systems the amount of capacity required for regulation and operating reserves is a small fraction of the capacity required to serve load. For example, in the PJM system, covering a large set of states centered in the U.S. Mid-Atlantic region, the regulation requirement is less than 1% of peak load. The amount of generation capacity required for primary reserves is less than 2% of peak load. The system has a category of secondary reserves that is much larger–more than 20% of peak load–but given that the capacity is installed, it is not costly to have it available with the speed defined for secondary reserves.[11]

Load Interruption and Load Shedding

As discussed in Chapter 2, certain customers may participate in direct load control programs, or may agree to have their service interrupted when necessary, which helps to provide stability and security to the system. This can be done on very short time scales using under-frequency relays which can interrupt power in fractions of a second or within seconds of detecting the drop in frequency. Or it can be done at longer time scales at the direction of the system operator to provide a form of reserve.

In some especially difficult cases, the system operator may not have sufficient reserves to match generation with load and may be forced to shed significant batches of loads without their agreement. This involuntary lost load is costly. Most systems monitor the amount of lost load as one gauge of system reliability.

4.3.2 Voltage Support

We have already discussed the importance of voltage and reactive power in connection with transmission capacity. While reactive power consumed by facilities or devices tends to cause the voltage to drop, it can also be produced or injected into the system to increase voltage in what is often referred to as **voltage support**. Along with forecasting load and developing a dispatch schedule, the system operator will develop a schedule of voltage support needed to maintain the system within its safe operating bounds. In addition, it will need to respond to changes that require extra voltage support.

---

[10] Riesz, Jenny, and Iain Macgill. "Frequency Control Ancillary Services Is Australia a Model Market for Renewable Integration?." (2013). Heffner, Grayson, et al. "Loads providing ancillary services: Review of international experience." Lawrence Berkeley National Laboratory (2008).

[11] Monitoring Analytics, LLC, Independent Market Monitor for PJM. State of the Market Report for PJM. Volume 1, p. 6, and Volume 2, pp. 390, 392, and 439.

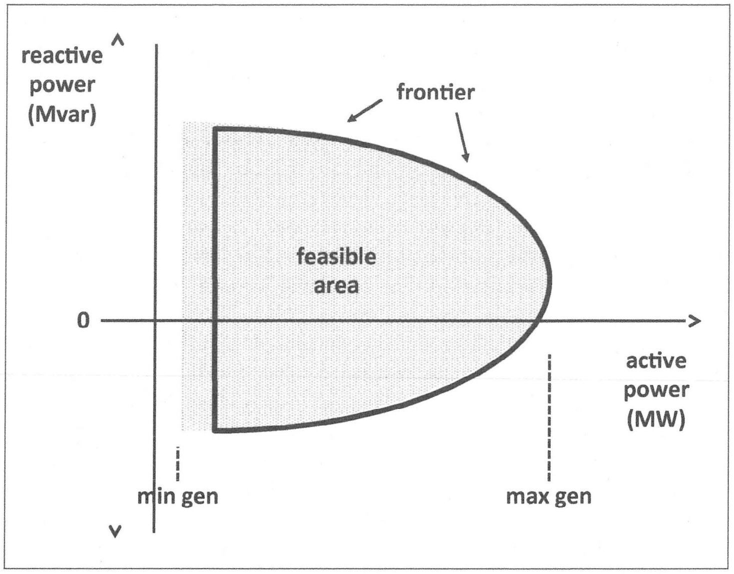


Figure 4.8: A Simplified Generation Capability Curve for a Single Generating Unit.

Certain generators can provide much of the needed voltage support by adjusting their reactive power output. Figure 4.8 is a 'generator capability curve', which shows the tradeoff between the generator's production of active power and the generator's production of reactive power. The horizontal axis is active power, measured in MW. The vertical axis is reactive power, measured in Mvar. The shaded area shows the feasible range of the generator's production of both active and reactive power. It can produce any combination within the area. The boundary of the area is the generator's capability frontier. The generator has a minimum level of generation of active power, which defines the leftmost boundary of the area in which the generator can operate, and it has a maximum level of generation of active power, which defines the rightmost point of the area. Note that the generator can either produce reactive power, or consume reactive power–the solid area straddles the line of zero reactive power production. Finally, it is important to note that the capability frontier at the top and the bottom shows a tradeoff between the production of active power and the range of feasible reactive power. As the generator's active power production ramps up, the maximum reactive power production declines and the minimum reactive power production increases. If the generator is operating inside the area, away from the boundary, the cost of providing extra reactive power is small. However, if the generator is operating at the upper boundary, on the frontier, then in order to provide extra reactive power, the generator will have to reduce its production of active power, which is an **opportunity cost**.

Reactive power can also be supplied by other power electronics equipment on the transmission system, including switched shunt capacitors, static var compensators, static synchronous compensators (also known as STATCOMs) and synchronous condensers.

4.3.3 Black Start

In addition to preventing outages, cascades and blackouts, it is important to plan for recovery. One key ingredient in a recovery plan is a set of generators which are capable of restarting on their own, without an external power source. In the U.S., this is known as **'black start'** capability. In Australia, it is known as system restart capability. Most large central station power plants need some external power source to supply a minimum amount of power to key functions in order for the main system to work. Therefore, in order to start without taking power from the grid, they need to have their own small power supply, such as

Table 4.4: A Sample of International Data on Losses as a Percent of Total Generation, 2001-2014.

Australia	6.7%	Hong Kong	13.0%	Poland	8.8%
Brazil	17.0%	India	24.2%	Russia	11.5%
Bulgaria	11.7%	Iran	17.3%	Singapore	3.6%
Canada	8.3%	Italy	7.4%	South Africa	9.1%
Chile	7.9%	Japan	4.8%	Spain	9.2%
China	6.6%	Jordan	13.7%	Taiwan	11.8%
Egypt	11.6%	Korea, South	3.8%	Turkey	11.8%
Finland	4.0%	Mexico	15.8%	United Kingdom	8.2%
France	6.3%	New Zealand	7.1%	United States	6.4%
Germany	4.7%	Nigeria	21.0%		

small back-up diesel generators. However, the system only needs a few of the plants to have this capability, and the amount of back-up needed is often less for certain plants than for others, so some plants provide this service while others do not. Black start services can also be provided by generators associated with specific isolated loads that, when sensing a system failure, can disconnect from the system and then reconnect and supply power to restart the larger system.

4.4 Losses

As power is transmitted across a transmission and distribution network, some of the power used up in the process, so that the total power delivered is less than the power generated. The difference is known as **losses**. These losses occur in various parts of the system, including magnetisation losses in transformers and the production of heat in transmission lines (a/k/a ohmic losses), among others. Therefore the total generation scheduled must equal the total load plus losses. The size of these losses depends upon characteristics of the transmission system, the paths over which the power flows, and the size of the power flows. Losses will also vary with other factors, such as temperature. Losses due to magnetisation are roughly fixed. Ohmic losses increase proportionally with the square of the power flow. Since ohmic losses vary inversely with voltage, long-distance transmission takes place using high-voltage lines. Ohmic losses on a line are roughly proportional to the length of the line, so systems in which generation is distant from load may have larger transmission losses, and systems in which individual customers are widely scattered may have larger distribution losses.

When dispatching generation, account must be taken to not only meet load, but also cover these losses. In addition, the total size of losses varies according to which generators are selected to serve load, since the choice of generators determines flows on the transmission lines. In later chapters we will see how losses affect the economic dispatch decision, and how losses enter into the market price of power at different locations.

Table 4.4 shows how large are transmission and distribution losses, as a percent of total generation, for a sample of countries. For many, losses account for 4-12% of total generation, although in some countries the number is larger.[12,13] A large fraction of losses occurs in the low voltage distribution network, with a smaller fraction occurring in the high voltage transmission network, although the allocation between the two varies by country characteristics. For example, in the PJM market, losses on the transmission system in 2014 are just over 2% of generation, which is approximately 1/3 of the total transmission and distribution

---

[12] The U.S. Energy Information Administration (EIA), International Statistics. https://www.eia.gov

[13] Two outliers in Table 4.4 are India and Nigeria, which report losses of 24.2% and 21.0%, respectively. Losses are calculated as measured generation minus measured consumption. Therefore, where some electricity is diverted off of the transmission system without being measured as consumption, reported losses will be larger that losses due to technical factors such as magnetisation of transformers and ohmic losses. The total losses in India and Nigeria may be especially large because this other portion of losses–also known as, non-technical losses–are material in these countries whereas non-technical losses are relatively small in most developed countries.

loss percentage in the U.S.[14]

4.5 Conclusion

Our focus in this chapter has been on select features of the transmission system. There are other features of the transmission system that we have not covered in more detail as they are outside the scope of this discussion. These include the different types of equipment that can be installed, different operational strategies for maintaining the transmission network and delivering transmission services, and itemizing the costs of component parts of the transmission network. Instead, we have taken the network as given, and we have focused on those features that shape the use of generation. Transmission limits constrain which generators can be dispatched to serve various loads. Ancillary services are provided by generators to help maintain system stability and reliability. In addition, losses are another important input to the generation schedule. As we shall see later, markets are only occasionally employed for transmission services, although they are often used to determine which generators provide power. Since this volume is about markets, we have chosen to focus on the aspects of the transmission system that shape the dispatch of generation.

It is important to keep clear how we are accounting for different costs. In Chapter 1, we mentioned the share of the total cost of delivering electricity that is associated with the transmission network. That should not be confused with the costs discussed in this chapter, which are the increased generation costs associated with transmission limits, the cost of providing ancillary services, and the cost of losses. These three costs are almost all charges for generation. So, the structure of the transmission network helps determine the size of generation charges. Increased investments in transmission will increase the transmission system's share of costs and decrease generation's share. Whether the investment lowers total costs is a complicated question we do not address here.

---

[14] Source: State of the Market Report for PJM for 2015, Volume 1, Table 1.

Study Review

Some key terms introduced in this chapter are:

- outage
- cascade
- blackout
- transmission capacity
- meshed network
- thermal limits
- voltage limits
- nominal voltage
- active power
- reactive power
- resistance
- reactance
- angle stability limits
- safety margin
- contingencies
- stability
- security
- N-1 criterion
- loop flows
- admittance
- ancillary services
- balance
- frequency control
- inertial response
- governor control
- direct control
- regulation
- operating reserves
- voltage support
- opportunity cost
- black start
- losses